J. Biochem. 128, 999–1008 (2000)

Sequence and Expression of Thai Rosewood β-Glucosidase/β-Fucosidase, a Family 1 Glycosyl Hydrolase Glycoprotein[1]

James R. Ketudat Cairns,*‡ Voraratt Champattanachai,*† Chantragan Srisomsap,* Brigitte Wittman-Liebold,¶ Bernd Thiede,§ and Jisnuson Svasti*†,2

Laboratory of Biochemistry, Chulabhorn Research Institute, Vipavadee-Rangsit Highway, Bangkok, 10210; †Department of Biochemistry, Faculty of Science, Mahidol University, Bangkok, 10400 Thailand; ‡Institute of Science, Suranaree University of Technology, Nakhon Ratchasima, 30000 Thailand; and ¶Max-Delbruck-Centrum für Molekulare Medizin, Abt. Proteinchemie, Robert-Roessle-Strasse 10, D-13125 Berlin-Buch, Germany

Received September 13, 2000; accepted September 29, 2000

Dalcochinin-8′-*O*-β-glucoside β-glucosidase (dalcochinase) from the Thai rosewood (*Dalbergia cochinchinensis* Pierre) has aglycone specificity for isoflavonoids and can hydrolyze both β-glucosides and β-fucosides. To determine its structure and evolutionary lineage, the sequence of the enzyme was determined by peptide sequencing followed by PCR cloning. The cDNA included a reading frame coding for 547 amino acids including a 23 amino acid propeptide and a 524 amino acid mature protein. The sequences determined at peptide level were found in the cDNA sequence, indicating the sequence obtained was indeed the dalcochinase enzyme. The mature enzyme is 60% identical to the cyanogenic β-glucosidase from white clover glycosyl hydrolase family 1, for which an X-ray crystal structure has been solved. Based on this homology, residues which may contribute to the different substrate specificities of the two enzymes were identified. Eight putative glycosylation sites were identified, and one was confirmed to be glycosylated by Edman degradation and mass spectrometry. The protein was expressed as a prepro-α-mating factor fusion in *Pichia pastoris*, and the activity of the secreted enzyme was characterized. The recombinant enzyme and the enzyme purified from seeds showed the same K_m for pNP-glucoside and pNP-fucoside, had the same ratio of V_{max} for these substrates, and similarly hydrolyzed the natural substrate, dalcochinin-8′-β-glucoside.

Key words: β-fucosidase, β-glucosidase, glycosylation, recombinant expression, plant.

β-Glucosidases [3.2.1.21] play important roles in plants, including growth regulation, response to stress, lignification, cellulose degradation, and defense (*1*). They are thought to regulate plant growth by releasing cytokinin growth factors from their glucosides, and they help hydrolyze cellobiose produced by cell wall degradation. β-Glucosidases involved in defense generally produce toxic compounds, such as hydrogen cyanide, saponins, coumarins, and naphthoquinones by deglycosylation of their substrates (*2–5*). Although these enzymes are closely related in evolution and carry out the same basic reaction, their substrate specificities are distinct (*2*, *6*). Prunasin hydrolase, for instance, is unable to hydrolyze amygdalin, linustatin, and neolinustatin (*3*, *7*). Other related β-glucosidases have evolved to fulfill diverse roles in plant growth (*8*), phosphate starvation response (*9*) and other plant processes, and are likely to have different aglycone specificities as well. Since these homologous enzymes show different substrate specificity, small sequence differences must affect substrate specificity.

The wide variety of β-glucosidases and other glycosidases in plants have several potential applications. Notably, they have been applied to synthesis of oligosaccharides by reverse hydrolysis (*10*) and mapping of carbohydrate linkages, including those in glycoproteins and glycolipids (*11*). For this reason, the discovery of new glycohydrolases with novel substrate specificities is of great interest.

During screening studies for glycohydrolases in Thai plants using *p*-nitrophenol (pNP) glycosides, seeds of Thai rosewood, *Dalbergia cochinchinensis* Pierre, were found to contain high levels of β-fucosidase and β-glucosidase activities (*12*). The pure enzyme had both β-glucosidase and β-fucosidase activities, an apparent subunit molecular weight of 66,000 by SDS-PAGE and an apparent native M_r of approximately 330,000 (*13*).

Both the apparent K_m and the k_{cat} for the enzyme were lower for the pNP-β-fucoside than the pNP-β-glucoside, while other pNP-glycosides had much lower k_{cat} values (*13*). As the k_{cat}/K_m was higher for the fucoside, the enzyme may be argued to be primarily a β-fucosidase. Modification of

[1] The work was supported by grants from the Chulabhorn Research Institute and the National Research Council of Thailand Thai-UK (NRCT-ROYSOC) program. JRKC is supported by grant RSA011/2539 from the Thailand Research Fund. Thai Rosewood dalcochinin 8′-*O*-β-glucoside β-glucosidase sequence accession: AF163097.

[2] To whom correspondence should be addressed at: Department of Biochemistry, Faculty of Science, Mahidol University, Bangkok 10400, Thailand. Fax: +66-2-2480375, 2596634, Tel: +66-2-2461360 (Ext 4305), E-mail: scjsv@mahidol.ac.th

Abbreviations: dalcochinase, dalcochinin 8′-*O*-β-glucoside β-glucosidase; endo F, endoglycosidase-F; endo Lys C, endoproteinase lysine-C; NCBI, National Center for Biotechnological Information; PAGE, polyacrylamide gel electrophoresis; PMSF, phenyl methyl sulfonyl fluoride; pNP, *para*-nitrophenol; RACE, rapid amplification of cDNA ends; RT-PCR, reverse-transcription and polymerase chain reaction.

the purified enzyme with conduritol β-epoxide (CBE) inactivated β-fucosidase and β-glucosidase activities at essentially the same rate, indicating the presence of a carboxylic acid in the common active site (*14*).

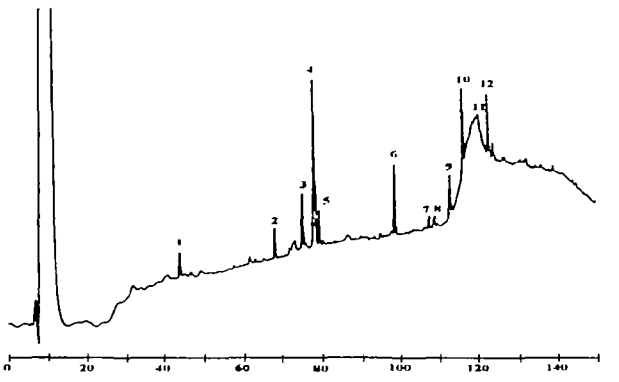
The enzyme does not efficiently hydrolyze commercially available natural glycosides, so the natural substrate was isolated from extracts of *D. cochinchinensis* seeds based on its cleavage by the enzyme (*13*, *15*). The natural substrate, which was present at over 3% (by weight) of the seeds, was found to be a novel isoflavanoid β-glucoside, which was named dalcochinin-8′-*O*-β-glucoside (*15*). Due to its structural similarity to rotenone, a natural insecticide and piscicide, the aglycone was postulated to play a role in plant defense against herbivores, such as insects. However, the molecule is distinct from other defense-related glycosides, and the respective β-glucosidases show little ability to hydrolyze each other's natural substrates (*15*).

Here, the primary structure for the *D. cochinchinensis* dalcochinin-8′-*O*-β-glucoside β-glucosidase (dalcochinase) was determined by peptide sequencing followed by RT-PCR amplification, cloning, and sequencing of the cDNA. The cDNA-derived protein sequence is shown to be closely related to known plant defense-related β-glucosidases in glycosyl hydrolase family 1 (*7*, *16*). To demonstrate that the clone represents the correct β-glucosidase isozyme, recombinant protein was expressed in *Pichia pastoris* and the resulting activity was characterized.

EXPERIMENTAL

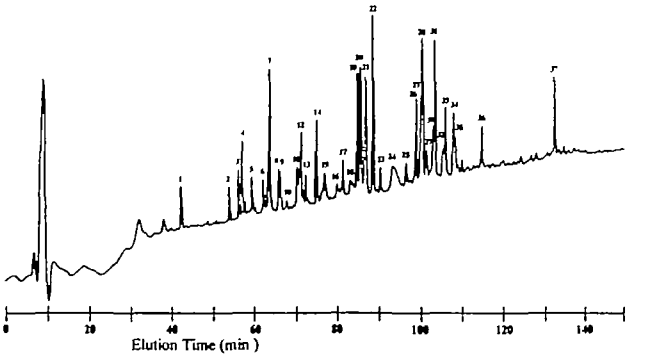
Materials and Reagents—D. cochinchinensis Pierre seeds were provided by the ASEAN-Canada Forest Seed Center at Muek Lek, Saraburi, Thailand. Endo-F/peptide-*N*-glycosidase F, pNP-glucose, and pNP-fucose were purchased from Sigma Chemical (St. Louis, MO, USA). Oligonucleotides were obtained from GIBCO-BRL (Life Technologies, Grand Island, NY, USA) and the BioServices Unit of the National Science and Technology Development Agency (NSTDA, BIOTECH, Bangkok, Thailand). Superscript II reverse transcriptase and the 3′ RACE kit were purchased from GIBCO-BRL. Trypsin, Taq polymerase, deoxyribonucleotides, other PCR reagents, pGEM-T cloning vector, and Wizard 373 plasmid preparation kit were purchased from Promega (Madison, WI, USA). Endopeptidase lysine-C, and terminal deoxynucleotide transferase were acquired from Boehringer Mannheim (Mannheim, Germany). A PRISM cycle sequencing kit from Perkin Elmer (Roche Molecular Systems, Branchburg, NJ, USA) was used in DNA-sequencing. *P. pastoris* strain GS115 and pIC9.5 plasmid were from Invitrogen (Carlsbad, CA, USA). All other reagents were of analytical grade or better.

Peptide Mapping and Protein Sequencing—The β-glucosidase/β-fucosidase enzyme was purified from *D. cochinchinensis* Pierre seeds as previously described (*13*). Deglycosylation was accomplished by incubating 0.5 mg of protein at 37°C for 24 h with 2.5 units of endoglycosidase-F/peptide-*N*-glycosidase F in 200 μl of 10 mM sodium phosphate (pH 7.5), 25 mM EDTA, and 0.01% sodium azide (final pH 6.7). Following deglycosylation, the protein was purified over an Aquapore RP-300 (220 × 4.6 mm) C8 column, eluted with a linear gradient from 0 to 70% acetonitrile in 0.1% TFA, on a Waters 510 HPLC. Portions of the enzyme preparation were digested with trypsin at an enzyme: substrate ratio of 1:50 by wt at 37°C in 50 mM Tris pH 7.85, 1 mM $CaCl_2$, 10% acetonitrile, or endoproteinase Lys-C (endo Lys C) at an enzyme: substrate ratio of 1:45 by weight in 50 mM Tris pH 9.0, 0.5 mM EDTA, 10% acetonitrile at 36°C for 12 h. The reactions were stopped by addition of 10% TFA to lower the pH to 2.0. The cleaved peptides were separated using a LichroCART 250–4 HPLC cartridge [Lichrospher 100 RP-8 encapped (250 × 4 mm)]



A Endo Lysine C Peptide Map and Determined Peptide Sequences.

Lys2:	I-D-F-A-K
Lys3:	Y-F-L-A-R-D-Q-X-S-A
Lys4:	I-A-D-X-S-N-G-D-V-A-V-D-Q-F-H
Lys5:	D-I-A-I-M-K
Lys6:	L-S-A-T-X-F-K
Lys9.	Y-N-N-P-L-V-Y-I-T-E-N-G-I
Lys10	G-S-F-X-F-I-G-L-N-Y-X-T-T-N



B Tryptic Peptide Map and Determined Peptide Sequences.

TP7	S-N-G-D-V-A-V-D-Q-F-H-Y-Y-N
TP7A:	S-N-G-D-V-A-V-D-Q-F-H-R
TP7B:	V-S-G-G-I-(N)-Q-T-G-V-D-Y-Y-N-(R)
TP9	D-M-N-L-D-A-Y
TP22	L-S-A-T-X-F-K
TP23	V-P-S-I-X-X-N-F-T
TP37	X-F-F-A-X-X-L-L-D

Fig. 1. **Proteolytic peptide maps of *Dalbergia cochinchinensis* Pierre β-glucosidase and sequences of peptides.** Purified and deglycosylated protein was digested with endopeptidase lysine C (A) and trypsin (B), and peptides were purified by reverse-phase HPLC as described in "EXPERIMENTAL." Peptides from selected peaks were sequenced by automated Edman degradation sequencing, and the determined peptide sequences are listed below the appropriate peptide map chromatograph. The peptide sequences are designated by their digestion enzyme (Lys for endopeptidase lysine C or TP for trypsin) and the peak number in the peptide map. Secondary inspection of the sequences of trypsin digest peak 7 indicated two signals, with the sequences TP7A and TP7B as indicated. Asparagine was not seen at position 6 of TP7B, but was derived from the nucleic acid sequence.

(Fig. 1). Purified peptide fractions were dried by speed vacuum.

From 200 pmol to 1 nmol of purified peptide fractions were dissolved in 50% acetonitrile in water, applied to a TFA-treated glass-fiber filter, dried under argon, and sequenced on an automated protein sequencer (Model 473A and 477 from ABI, Foster City, CA, USA). Some fractions were also studied by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF) on a VG TOFSpec mass spectrometer (Fisons Instruments, Manchester, UK). A 0.8-μl sample was mixed with 1.2 μl of matrix: a 1:1 mixture of saturated α-cyano-4-hydroxy cinnamic acid in acetonitrile/0.1% aqueous trifluoroacetic acid, then applied to the probe, ionized with a 337 nm nitrogen laser, and accelerated with a voltage of 22 kV.

Determination of Enzyme Activities in Thai Rosewood Tissues—Seeds were surface-sterilized in 0.1% bleach, soaked in distilled water overnight, and germinated on moist cheese cloth. Young leaves and stems were collected from *D. cochinchinensis* Pierre trees and extracted within 1 h of collection. Tissue samples (0.3–0.5 g) were weighed, quick frozen in liquid nitrogen, ground in a cold mortar and pestle, mixed with 2 ml/g tissue of cold 0.05 M sodium acetate, pH 5.0, 1 mM PMSF, and 5% (w/v) polyvinyl polypyrrolidine (PVPP), and filtered through four layers of cheese cloth. The homogenate was then centrifuged at 8,940 ×*g* (4°C) for 30 min. Dowex 2X-8 resin was added to the supernatant up to 25% (w/v), stirred for 1 h at 4°C, and removed by centrifugation at 8,940 ×*g* for 30 min. The supernatant was either assayed immediately or stored at −20°C until assay. Activities toward pNP-β-glucoside and pNP-β-fucoside were determined as previously described (*12*), using a Shimadzu UV-2100S spectrophotometer (Shimadzu, Tokyo) to measure the *p*-nitrophenol absorbance at 400 nm.

PCR Amplification and cDNA Cloning—The general cloning and sequencing strategy and cloned products are outlined in Fig. 2. Degenerate oligonucleotide primers were designed from the N-terminal sequence and the sequences of two internal peptides (TP7 and Lys9). The N-terminal sequence was back-translated to the For.1 primer, while the TP7 peptide sequence was converted to Rev.1 primer, and the Lys9 peptide sequence to the Rev.2 primer, shown in Table I.

D. cochinchinensis seedlings were ground in liquid nitrogen and RNA was extracted with Trizol reagent, according to the recommended procedure (*17*). The RNA obtained was dissolved in 0.100 ml of DEPC-treated water, and 1–5 μg of total RNA was reverse-transcribed into single strand cDNA using Superscript II reverse transcriptase, with 2.5 μM Rev.2 primer (50 pmol per 20 μl), as described by the manufacturer. A 1-μl aliquot of the product was used as template for polymerase chain reaction (PCR) with 1 μM each of For.1 and Rev.1 primers, 1.5 mM $MgCl_2$, 50 mM KCl, 0.1% Triton X-100, 0.2 mM of each dideoxy-nucleotide (dATP, dCTP, dGTP, and dTTP) and 2.5 units of Taq polymerase in a 50-μl reaction. All PCR reactions were done in a Perkin-Elmer 480 thermocycler using a 94°C melting temperature and 72°C as extension temperature and were begun with 5 min melting and ended with 7 min extension. For the initial PCR , 5 cycles of 94°C for 30 s, 37°C for 30 s, and 72°C for 60 s were followed by 30 cycles of 94°C for 30 s, 44°C for 30 s, and 72°C for 60 s.

Four oligonucleotide primers, For.2, For.3, Rev.3, and Rev.4, were made from the sequence of the intitial PCR product for use in rapid amplification of cDNA ends (RACE). A 3′ RACE system kit from GIBCO-BRL was used for amplification of the 3′ end of the cDNA. The first-strand cDNA was produced as described above, but 10 pmol of AP (poly T) primer provided in the kit was used for priming the synthesis. The RNA was degraded by incubating with RNase H for 20 min at 37°C. The cDNA was first amplified with 0.25 μM each of For.2 and AUAP primers (Table I) in a 50-μl reaction, containing the other reagents as described above. The amplification was done by 35 cycles of 94°C for 1 min, 60°C for 1 min, and 72°C for 3 min. A second amplification used 5.0 μl of the first reaction as template and 10 pmol For.3 and AUAP primer, under the same conditions as the first amplification. The product was evaluated by 1% agarose gel electrophoresis (*18*) and gel-purified using the Geneclean III (Bio 101 Co., Vista, CA, USA) method.

Similarly, the 5′ end of the cDNA was amplified by 5′ RACE. Synthesis of the first strand cDNA was primed using 1 μM Rev.2 primer under the same conditions as 3′ RACE. After heat inactivation of the reverse transcriptase, the RNA was digested by adding 1 unit of RNase H to the reaction and incubating at 37°C for 20 min. The first-strand cDNA was then purified with Geneclean III and eluted in 10 μl of sterile, distilled water. The product was tailed with dGTP using 25 units of terminal deoxy-transferase (TdT) in 20 μl of 0.5 M potassium cacodylate, 25 mM Tris-HCl, pH 6.6, 0.25 mg/ml BSA, 0.1 mM DTT, 1 mM $CoCl_2$, and 25 μM dGTP incubated at 37°C for 20 min. After heat inactivation at 70°C for 15 min, 10% of the product was amplified by PCR with 1 μM each of Oligo dC14-*Sal*I and Rev.3 primers and 5 cycles of 1-min incubation with annealing at 37°C,

TABLE I. **Oligonucleotides used in cloning and sequencing.**

A. Degenerate oligonucleotides used for RT-PCR
For.1 5′-GGGATCC AT(T/C/A) GA(T/C) TT(T/C) GC(N) AA(A/G) GA(A/G) GT-3′
Rev.1 5′-GGGAAGC TT (A/G)TA (A/G)TA (A/G)TG (A/G)AA (C/T)TG(G/A)TC-3′
Rev.2 5′-CC(A/G) TT(C/T) TC(N) GT(A/G/T) AT(A/G) TA-3′
B. Oligonucleotides used in 3′ RACE
AP 5′-GGC CAC GCG TCG ACT AGT ACT TTT TTT TTT TTT TTT T-3′
AUAP 5′-GGC CAC GCG TCG ACT AGT AC-3′
For.2 5′-AGG TTC CTC CAT TCA ACC GAA G-3′
For.3 5′-CCA CCA ATA TCC AGA AAA GAT AGC G-3′
C. Oligonucleotides used in 5′ RACE
Oligo-dC-SalI 5′-TTG TCG ACC CCC CCC CCC CCC-3′
Rev.3. 5′-AAC ATC TCC GTT GCT TCT ATC CGC-3′
Rev.4. 5′-GCT GTC CCA AAA ATG AAA TCT GA-3′
D. Oligonucleotides used for sequencing 3′ RACE product and full cDNA
For4. 5′-AGC CTT GGA GGA TGA GTA CGG T-3′
For5. 5′-TCA AAA TGC TAC CCA GCG ATA TCT-3′
For6. 5′-CGC CAT CTC TTT TAT ATT CGA TAT GC-3′
For7. 5′-GGC AAG ATG GAG CTT ATC AAC G-3′
Rev5. 5′-TTT TGT CTT TTC CTT GAT CAT CG-3′
Rev6. 5′-CGC CAG ACC TAA TTG CAT ATC G-3′
Rev7. 5′-CAT CCA CAT GTG AAG TCA AGA TAT CG-3′
Rev8. 5′-AAT TGG AAF CAA AGA TCC GCA TA-3′
E. Oligonucleotides used to amplify full coding region of cDNA
5′-Term 5′-CTT CCT TTC ATC TCA TGC TTG CA-3′
3′-Term 5′-AAA GAG AAT ACA ATT CTT TTT GGG CG-3′
F. Oligonucleotides used for cloning into pIC9K
9KF1 5′-TT AGT TAC GTA ATT GAC TTT GCA AAA GAA GTC CGT GA-3′
9KR1 5′-TT ATT CCT AGG CAA TCA AAA GCC TTC AAT GC-3′

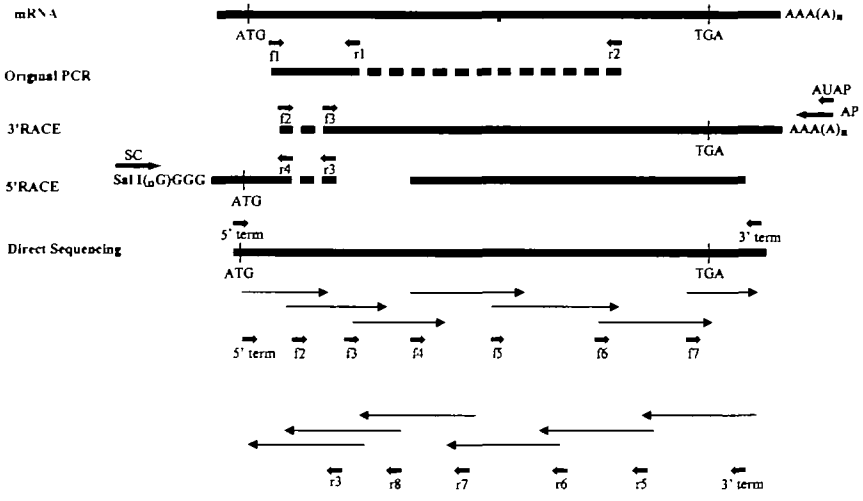


Fig. 2. **Map of the sequence clones and peptide sequences of the β-glucosidase.** The sequence of the entire mRNA (cDNA) is represented by the top band marked mRNA. The primers (small arrows) used for initial reverse transcription and PCR are indicated below this band, followed by the band representing the original PCR product. Below this are the primers designed from this sequence and the other primers used for 3′ and 5′, followed by 3′ RACE and 5′ RACE products. The 5′ term and 3′ term primers used for amplification of the full coding region for sequencing and their amplification product are shown below this. Other primers used for sequencing are shown below this product, along with segments sequenced by direct sequencing (indicated by long, thin arrows). These primers were also used for sequencing the cloned RACE products.

followed by 30 cycles of annealing at 60°C. A second amplification was performed using 5 μl of the first reaction and Oligo dC14-*Sal*I and Rev.4 primers under the same conditions as in 3′RACE.

All PCR products were ligated into pGEM-T vector and transformed into *Eschericia coli* DH5-α according the supplier's instructions. Mini-preparations of the cloned products were prepared using the Wizard 373 plasmid preparation kit. The DNA was digested with *Pst*I and electrophoresed in 1% agarose to verify inserts (*18*). The inserts were sequenced by automated sequencing using M13 forward and reverse primers and internal primers derived from the cloned sequences (Table I) on an ABI 373A DNA sequencer.

The full-length cDNA was amplified using the 5′-term and 3′-term primers (Table I). The product was gel-purified using the Geneclean III procedure as above and sequenced directly using the internal primers (Fig. 2 and Table I). All sequences were determined from products of at least two independent PCR reactions.

Sequence Analysis—Sequences were compared to the nr database for related sequences, using the BLAST facility of the National Center for Biotechnology Information at the USA National Institute of Health to identify related proteins (*19*). The translated protein sequence was aligned with related enzyme amino acid sequences using the programs of Feng and Doolittle with the Dayhoff 250 and Blosum 62 similarity matrices (*20*). Structural comparisons between the Thai rosewood enzyme and the cyanogenic β-glucosidase from white clover (1CBG) (*21*) were made using the SwissProt PDB viewer and SWISSMODEL facility (*22*).

Protein Expression in P. pastoris—The cloned β-glucosidase cDNA was amplified with the primers listed in Table IF to introduce a *Sna*B site at the amino terminal of the mature protein and an *Avr*II at the 3′ end of the coding region, using Pfu polymerase, and sequenced to confirm sequence integrity. The cDNA was then digested and cloned into the *Sna*B and *Avr*II sites of the pIC9K plasmid according to standard methods. The plasmid was transformed into *P. pastoris* strain GS115 according to the supplier's procedure. Transformants were grown on histidine-free media to select those integrating the DNA. Selected clones were then grown in BMMY medium, and cells were separated from media by centrifugation (4°C). Cells were lysed by vortexing with glass beads, and soluble and insoluble fractions were separated by centrifugation (4°C) according to the supplier's directions (Invitrogen). The media, soluble and insoluble cell lysates were tested for β-glucosidase expression by pNP-β-glucoside and pNP-β-fucoside hydrolysis (*12*). Clone 16, expressing β-glucosidase activity, was grown in 200 ml of BMM supplemented with methanol to induce protein production, according to the recommended method (Invitrogen), and the culture was harvested after 3 days. The cells were removed by centrifugation, and the medium was supplemented with 1 mM PMSF and 1 mM EDTA and concentrated to 7 ml by ultrafiltration (10,000 M_rCO) at 4°C. The medium was then washed six times with 100 ml of 1 mM PMSF, 1 mM EDTA to remove medium components. Concentrated medium was characterized by non-denaturing gels and stained for activity with MU-glucoside as previously described (*13*). Kinetic parameters (K_m and V_{max}) were determined using pNP-glucoside, pNP-fucoside, and dalcochinin-8′-*O*-β-glucoside as substrates by pNP and glucose production as previously described (*13*, *15*).

Thin Layer Chromatography—Thin layer chromatography (TLC) was performed to visualize products of pNP-β-glucoside and dalcochinin 8′-*O*-β-glucoside hydrolysis using 0.2 mm C18 reversed-phase gel (RP-18) from Merck. Portions of 5 mU of concentrated media, purified enzyme from seed, cassava linamarase, or almond β-glucosidase were incubated with 2 mM final concentration of dalcochinin 8′-*O*-β-glucoside and pNP-Glc in 100 mM sodium acetate, pH 5.0 at 37°C for 1 h. Reaction mixtures (7 μl) were spotted on the TLC plate and chromatographed vertically using acetic:methanol:water (1:4:5 by volume) mixture for 30 min. The plate was sprayed with concentrated H_2SO_4 and baked at 105°C for 15 min to visualize the reaction products.

RESULTS AND DISCUSSION

Protein Sequencing and Analysis—The β-fucosidase/β-glucosidase purified from seeds of *D. cochinchinensis* Pierre was shown to be homogeneous by SDS-PAGE and native gel electrophoresis (*13*). The protein was deglycosylated with endo-F/endoglycosidase F, which decreased its apparent M_r on SDS-PAGE from 66,000 to 63,000. After purification, the deglycosylated protein was sequenced by auto-

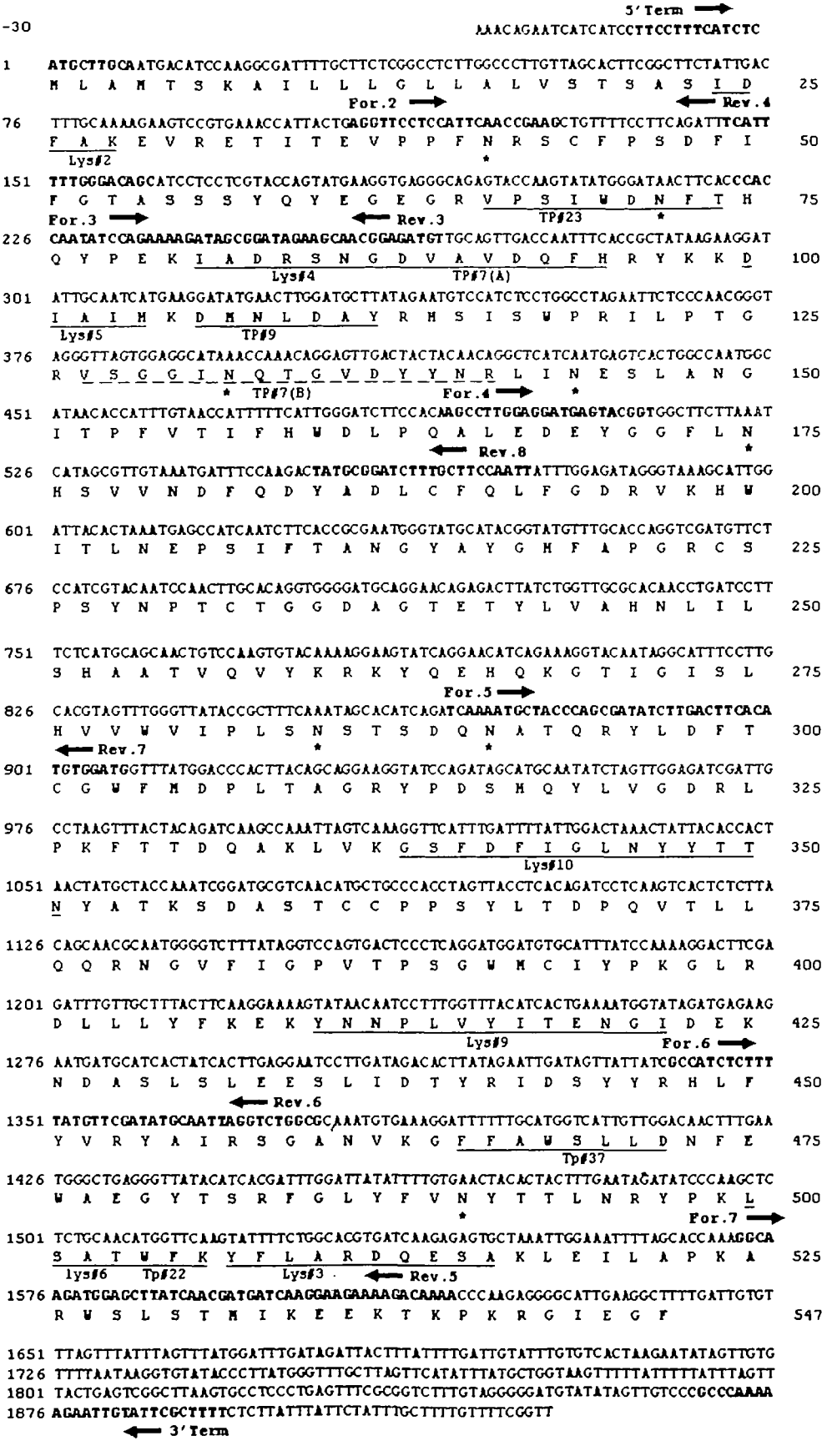


Fig. 3. **Nucleic acid and protein sequence of the *D. cochinchinensis* β-glucosidase, Genbank accession, AF-163097.** Peptides are underlined and labeled under the sequences. Possible glycosylation sites are marked by stars and the site confirmed by mass-spectroscopy and peptide sequencing is in bold. Primer sequences are indicated by bold print in the nucleotide sequence.

mated Edman degradation to yield the N-terminal sequence: IDFAKEVR. Endopeptidase lysine-C and trypsin peptides were separated by HPLC as shown in Fig. 1, A and B, respectively. The sequences determined by Edman degradation are indicated beneath the peptide maps of Fig. 1.

To determine the best tissue from which to extract RNA for PCR cloning, crude enzyme was extracted from several tissues and the relative amounts of β-glucosidase and β-fucosidase activities were determined. Both dried seeds, from which the enzyme was originally extracted, and germinated seeds had greater than 100 times the level of β-glucosidase activity found in leaf, stem, and shoot tip and approx. 450 times the levels of β-fucosidase activity found in these tissues. Dried and germinated seeds each had a 2.2-fold ratio of activity toward 1 mM pNP-β-D-fucoside compared to activity toward 1 mM pNP-β-D-glucoside, in agreement with the previously characterized enzyme purified from dried seeds (*12*, *13*). Leaves, stems, and shoots, in contrast, had much lower levels of β-fucosidase relative to β-glucosidase, indicating that other β-glucosidases may predominate in these tissues. Germinated seed had twice the activity per seed, indicating that the enzyme was newly synthesized or activated during germination. Therefore, germinated seeds were used to prepare RNA for RT-PCR cloning of the β-glucosidase cDNA.

RT-PCR of RNA prepared from germinated *D. cochinchinensis* seeds amplified with the For.1 and Rev.1 primers designed from the N-terminal and TP7 peptide sequences produced a specific product. The sequence of this cloned product contained 224 bp, including the primer sequences, which translated to 74 amino acids at the amino terminal of the mature β-glucosidase sequence. The derived protein sequence included the Lys4 and TP23 peptide sequences in addition to the N-terminal and TP7 sequences in the For.1 and Rev.1 primers (Fig. 3). RACE yielded the sequences of the 5′ and 3′ ends of the cDNA, which indicated a total cDNA length of 1,906 nucleotides, most of which was sequenced directly to eliminate errors in PCR amplification (Fig. 3, Genbank accession AF163097) (*23*).

This sequence included an uninterrupted reading frame coding for 547 amino acids, starting from the first AUG Met codon and ending with a TGA stop codon. It also included the *N*-terminus of the mature protein and a 23 amino acid pre-sequence, starting from the first methionine (Fig. 3). The pre-peptide included a typical signal sequence for transport into the endoplasmic reticulum. After the stop codon, there were 292 nucleotides of putative 3′ untranslated region before the polyA tail, corresponding to the reverse transcription primer.

Analysis of Peptide Sequence—The sequence also included the sequences of the Lys2, Lys3, Lys4, Lys5, Lys6, Lys9, Lys10, TP9, TP22, and TP37 peptides (see Fig. 3). However, some nucleotides in the Rev.1 primer used to amplify the initial PCR fragment did not match the new sequence. Notably, the amino acid sequence VDQFHRYK was found, instead of the VDQFHYYN of the TP7 peptide sequence used to design the primer. Inspection of the translated cDNA sequence revealed another predicted tryptic peptide, starting from the valine at residue 127, with the sequence YYN in positions 12 to 14 (as in TP7). Closer inspection of the protein sequencing chromatographs indicated minor peaks in cycles 1–11 consistent with the sequence VSGGINQTGVDYYN (peptide T7B in Figs. 1 and 3), and a small arginine peak at cycle 12 (consistent with the TP7A sequence of Figs. 1 and 3). However, no asparagine peak was seen in cycle 6.

MALDI-TOF mass spectrometry of the HPLC peak fraction (Fig. 4) indicated two peaks at 1,345.5 amu, consistent with the sequence SNGDVAVDQFHR (TP7A), and 1,846.9 amu, consistent with VSGGINQTGVDYYN (TP7B) plus one *N*-acetyl-glucosamine. That the protein was deglycosylated by endoglycosidase F, which leaves one *N*-acetyl-glucosamine residue, and that the Edman degradation did not provide the predicted Asn peak at cycle 6, suggest that Asn109 is indeed a site of *N*-linked glycosylation.

Sequence Comparison—BLAST search of the NCBI nr database showed that the protein was most closely related to glycosyl hydrolase family 1 plant β-glucosidases (*7*, *16*). Phylogenetic analysis indicated that dalcochinin 8′-*O*-glucoside β-glucosidase is clustered with the dicotyledon defense-related β-glucosidases, such as linamarase and amygdalin and prunasin hydrolases. The mature protein showed highest identities with the white clover cyanogenic and noncyanogenic β-glucosidases (60 and 56%, respectively) and the black cherry amygdalin hydrolase isoform AH I (56%). Due to the high sequence identity with the cyanogenic β-glucosidase from white clover, the dalcochinase is expected to have an $(\alpha/\beta)_8$ barrel structure, very similar to the crystallographic model for that protein (1CBG) (*21*).

The alignment (Fig. 5) showed that the glutamic acids previously identified as the catalytic nucleophile and the catalytic acid are conserved in the *D. cochinchinensis* protein at positions E 396 and E 182, respectively, of the mature protein (E 419 and E 205 of the precursor protein). Glutamate 396 was postulated to be the catalytic nucleophile due to its homology to Glu358 in the β-glucosidase from *Agrobacterium*, which has been shown to act as the catalytic nucleophile (*24*). Similarly, Glu182 was homologous to Glu198 of the cassava cyanogenic β-glucosidase, which was shown to be the catalytic proton donor in that

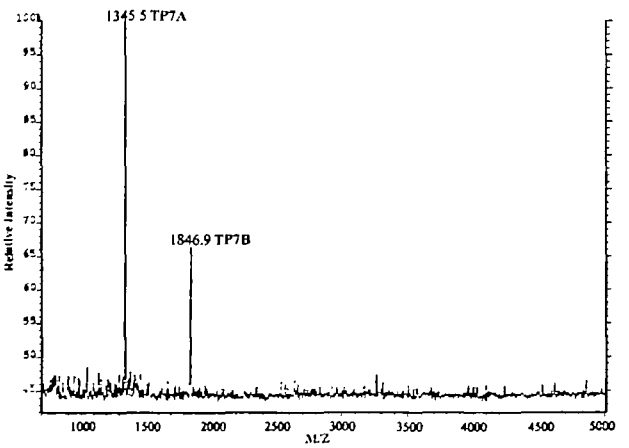


Fig. 4. **Mass spectrum of the tryptic digest HPLC fraction TP7.** The fraction was isolated by HPLC and the mass spectrometry conducted by matrix-assisted laser desorption/ionization with time-of-flight mass analysis (MALDI-TOF). The two major peaks are labeled with their molecular weights (m/z with $z = 1$) and associated peptide sequence. Predicted masses are: TP7A (SNGDVAVDQFHR), 1,344.6 amu; and TP7B (VSGGINQTGVDYYNR with *N*-acetylglucosamine), 1,845.9 amu.

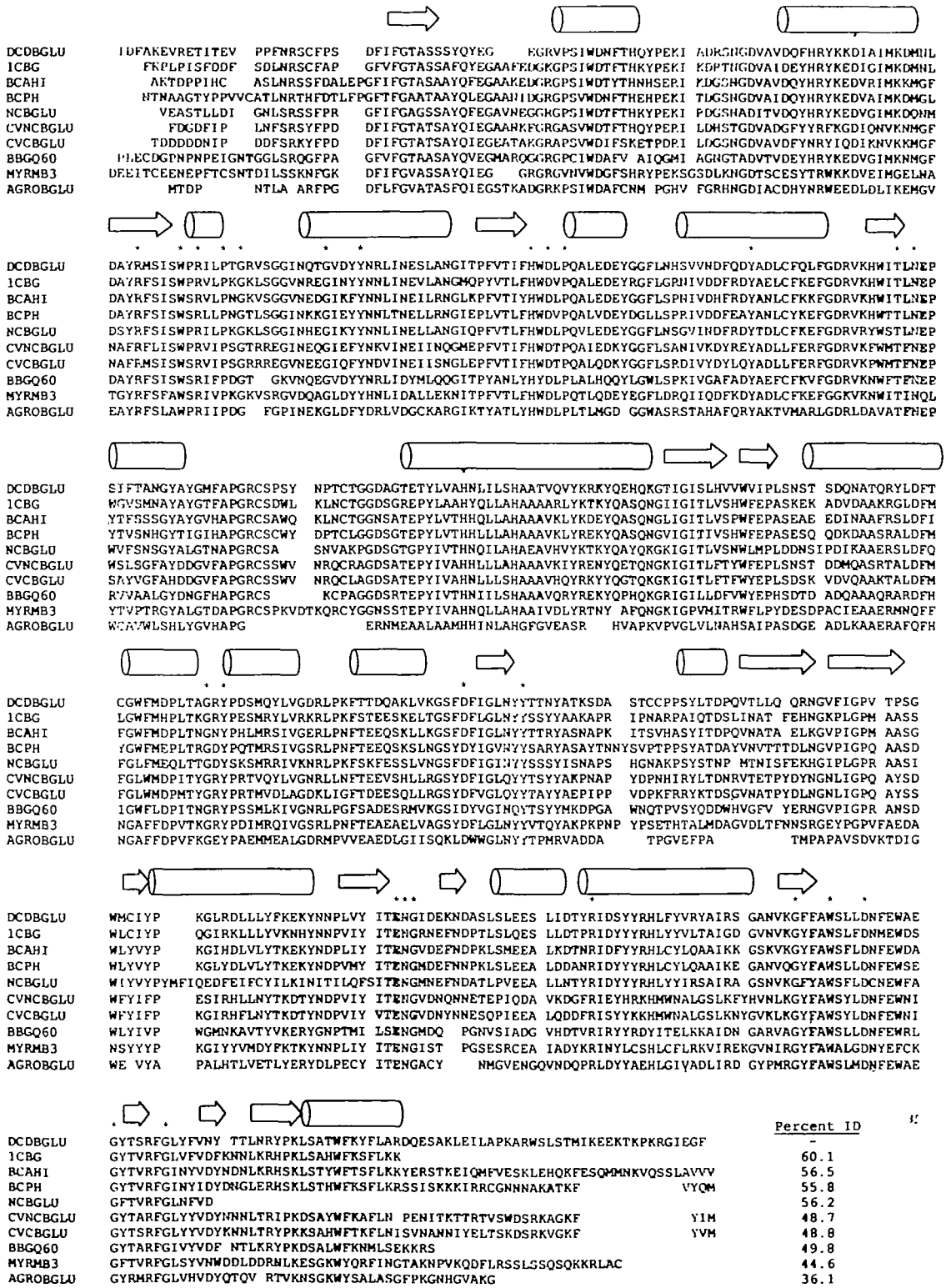


Fig. 5. **Multiple sequence alignment of *D. cochinchinensis* β-glucosidase with related enzymes.** The sequence of the mature protein for the *D. cochinchinensis* β-glucosidase is aligned with those of closely related mature proteins, plus white mustard myrosinase and *Agrobacterium* β-glucosidase, in which the catalytic nucleophile was originally identified. Stars indicate residues that are completely conserved among the sequences. The catalytic nucleophile (Glu395) and catalytic acid/base (Glu182) are indicated in bold font and other residues found within 6 Å of these residues in the cyanogenic β-glucosidase (1CBG) crystal structure are shaded. The secondary structure derived from the 1CBG X-ray crystal model is diagrammed above the sequences as arrows for β-strands and cylinders for α-helices. The sequences (and nr accessions) are: DCDBGLU, *D. cochinchinensis* β-glucosidase (dalcochinase, AF163097); 1CBG, white clover cyanogenic β-glucosidase (1311386); BCAHI, black cherry amygdalin hydrolase AH-1(833835); BCPH, black cherry prunasin hydrolase (1236961); NCBGLU, white clover non-cyanogenic β-glucosidase, (114974); CVNCBGLU, cassava non-cyanogenic β-glucosidase (1155090); CVNCBGLU, cassava cyanogenic β-glucosidase (linamarase) (249262); BBGQ60, barley β-glucosidase BGQ60 (136216); MYRMB3, white mustard myrosinase 3D (127734), *Agrobacterium* sp. (strain ATCC 21400) β-glucosidase (114966).

enzyme (*25*). Though these exact positions have not been shown experimentally, the carboxylic acid residues are consistent with the inhibition of the enzyme by mercuric compounds and CBE (*13*, *14*).

Other residues reported to be near the active site of the white clover cyanogenic β-glucosidase were largely conserved in the dalcochinase (Fig. 5). Of the residues noted by Barrett *et al.* (*21*) to surround the active site, only Val254 of the white clover enzyme is replaced (by His253) in the dalcochinase. This change may result in a less hydrophobic environment for Glu182, which was proposed to be in a hydrophobic pocket in the 1CBG structure. Ser184, which is Trp in the 1CBG structure, is also predicted to be within 6 Å of the active site, based on homology. These residues lie between the catalytic residues and the protein surface in the cleft where the aglycone is likely to bind, so they might also influence the substrate specificity. A substitution of Ser for Trp may allow the large rotenoid structure of dalcochinin to fit into this cleft more easily and interact with the many aromatic residues in this region. The smaller structures of linamarin and lotaustralin, the substrates for the white clover enzyme, may allow the Trp to fit in this position, although cassava linamarase has Ser at this position as well (Fig. 5). More structural and functional studies will be necessary to elucidate which amino acids contact the aglycone in the entrance to the catalytic pocket and account for the differences in specificity between these enzymes.

The alignment also showed that all the cysteines found in the 1CBG structure are conserved, and additional cysteines are found at positions 278, 338, and 339 in the mature dalcochinase. The significance of the conservation is unclear, since only one pair is predicted to form an internal disulfide, but these cysteines are near the surface in 1CBG, and interchain disulfides may stabilize the quaternary structure. Indeed, protein of higher M_r is found on SDS-PAGE when no reducing agent was added (result not shown).

An increase in electrophoretic mobility on SDS-PAGE after treatment with endo F (from an apparent M_r of 66,000 to 63,000) together with the observed *N*-linked sugar in the TP7B peptide indicated that dalcochinase is a glycoprotein, which passes through the endoplasmic reticulum and is *N*-glycosylated. This, and the presence of a signal sequence in the prepeptide, suggest that the protein may be localized in the secretory organelles, vacuole, or outside the cells. Other β-glucosidases have been found in the cytoplasm and plastids as well as in the lysozome and associated with the cell wall (*1–4*, *26*, *27*). It will be interesting to further characterize the cellular location of the enzyme and its substrate, since many defense-related β-glucosidases are found in separate compartments from their substrates.

A total of eight possible sites of *N*-linked glycosylation were identified in the sequence (Fig. 3). This is the same number of sites as are seen in the white mustard myrosinase crystal structure and significantly more than the four possible sites seen in the 1CBG protein (*28*). As both dalcochinase and white mustard myrosinase are found in seeds, while 1CBG is found in white clover leaves, this supports the suggestion of Burmeister *et al.* (*28*) that the high level of glycosylation may support solubility in the relatively dry seed environment.

The evolutionary relationship with β-glucosidases that generate noxious compounds suggests that the dalcochinase enzyme may also play some role in defense against herbivores, as does its substrate, which is similar to the respiratory inhibitor rotenone. This role is being investigated by studies of the bioactivity of the substrate. The apparent specificity of expression of enzyme activity in seeds and young plants will also need to be explored further to determine the biological significance of the enzyme. β-Glucosidases have also been implicated in activation of cytokinin growth factors and phosphate stress response (*8*, *9*). So, it is not yet possible to rule out the possibility that the enzyme and its substrate have a role in growth or environmental interactions apart from defense against herbivores.

Recombinant Protein Expression and Characterization—To test whether the cloned gene had the same catalytic properties as the enzyme purified from the plant, the recombinant protein was expressed from the cloned cDNA. Attempts to express the protein in *E. coli* resulted in only insoluble, inactive protein, so the protein was expressed in the yeast *P. pastoris* by fusing the mature protein coding region of the *D. cochinchinesis* β-glucosidase to the *Saccharomycees cervisiae* alpha mating factor prepropeptide. As shown in Table II, upon induction, β-glucosidase activity was detected both inside the cell and in the medium. A *Pichia* β-glucosidase was also found inside the cell, however, the *P. pastoris* enzyme did not hydrolyse pNP-fucoside and was not found in the medium. Recombinant β-glucosidase was concentrated from the medium and shown to hydrolyze pNP-glucoside and dalcochinin β-glucoside, as did the enzyme from seed, but control medium concentrate could not hydrolyze these substrates (Fig. 6). Almond β-glucosidase and cassava linamarase did not detectably hydrolyze dalcochinin β-glucoside, demonstrating the unique substrate specificity of the Thai rosewood β-glucosidase, which was exhibited by the recombinant enzyme.

TABLE II. **β-Glucosidase and β-fucosidase expressed in *Pichia pastoris*.**

	β-Glucosidase/ml culture	β-Fucosidase/ml culture
Recombinant cell extract	0.048 U	0.102 U
Recombinant medium	0.077 U	0.186 U
Nonrecombinant cell extract	0.003 U	N.D.[a]
Nonrecombinant medium	N.D.[a]	N.D.[a]

[a]N.D. means not detectable.

TABLE III. **Kinetic parameters for recombinant and seed-derived enzyme.**

Substrate	K_m (mM)		V_{max} relative to pNP-glucoside	
	Recombinant	Seed[a]	Recombinant	Seed[a]
pNP-β-D-glucoside	5.78 ± 0.20	5.37 ± 0.09	1	1
pNP-β-D-fucoside	0.52 ± 0.06	0.54 ± 0.04	0.48	0.49
Dalcochinin 8′-*O*-β-glucoside	1.08 ± 0.08	1.68 ± 0.11	1.34	1.06

[a]Values for seed-purified enzyme are taken from Refs. *13* and *15*.

To further characterize this activity, the K_m values for hydrolysis of pNP-glucoside, pNP-fucoside, and dalcochinin β-glucoside by the recombinant β-glucosidase was determined. As shown in Table II, the recombinant enzyme had the same K_m for pNP-glucoside and pNP-fucoside and a slightly lower K_m for dalcochinin β-glucoside (1.08 mM *vs.* 1.68 mM) compared to the β-glucosidase purified from seeds. The ratio of V_{max} for pNP-glucoside compared to pNP-fucoside was 2:1 for the recombinant enzyme, very similar to that of the enzyme from the seed. So, the catalytic properties of the recombinant protein are almost indistinguishable from those of the seed-derived protein. The difference in K_m for the natural substrate is small and may be due to differences in the batch of substrate, which must be weighed for the assay. Different batches may have different water contents, which could affect the determination. However, the difference could also reflect differences in post-translational modification, such as glycosylation, in *P. pastoris* compared to the plant seed. In fact, the recombinant protein gave a very broad band of activity staining on a nondenaturing gel (data not shown), which may be due to variable glycosylation and possibly proteolytic processing (*29*).

To our knowledge, this is the first report of the production of a plant β-glucosidase in *P. pastoris* (*29*). Previously, cassava linamarase was produced as a fusion protein with glutathione-S-transferase (GST) at low activity levels (*30*) and the unglycosylated maize chloroplast β-glucosidases have been expressed as free proteins at relatively high levels in *E. coli* (*31*). However, many characterized plant β-glucosidases are glycoproteins that may not express well in bacterial systems. So, *P. pastoris* and other yeast systems may provide a better expression system for these proteins. However, the presence of an endogenous β-glucosidase activity in *P. pastoris* suggests secretion systems such as ours may work better than expression of the protein within the cytoplasm. The amount of expression is low [1.6 mg/liter in minimal medium (BMM) or 5.2 mg/liter in rich medium (BMMY), assuming the enzyme has the same specific activity as that purified from seeds]. However, this is not unusual for initial expression of proteins in *P. pastoris* shake-flask cultures, and the expression system can be further optimized for high expression (*32*).

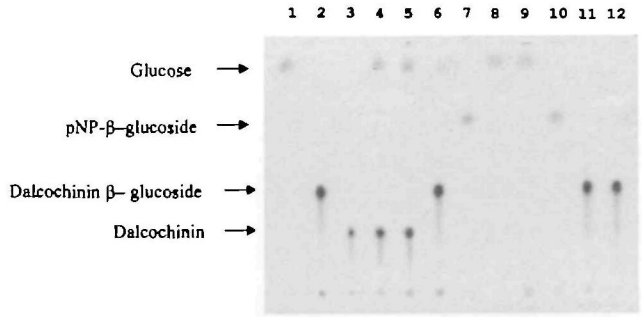


Fig. 6. **TLC of products of dalcochinin β-glucoside and pNP-β-glucoside hydrolysis by recombinant and pure β-glucosidases.** Substrates (2 mM) were incubated with enzyme, and samples of 7 μl were loaded onto reverse-phase TLC and developed as described in "EXPERIMENTAL." Lanes are as follows: 1, glucose standard; 2, dalcochinin 8′-*O*-β-glucoside standard; 3, dalcochinin standard; 4, reaction of seed dalcochinase with dalcochinin 8′-*O*-β-glucoside; 5, reaction of recombinant 16/21 medium with dalcochinin 8′-*O*-β-glucoside; 6, reaction of non-recombinant GS115 *P. pastoris* medium with dalcochinin 8′-*O*-β-glucoside; 7, pNP-glucoside standard; 8, reaction of seed dalcochinase with pNP-glucoside; 9, reaction of recombinant 16/21 *P. pastoris* medium with pNP-glucoside; 10, reaction of non-recombinant GS115 *P. pastoris* medium with pNP-glucoside; 11, reaction of cassava linamarase with dalcochinin 8′-*O*-β-glucoside; 12, reaction of almond β-glucosidase with dalcochinin 8′-*O*-β-glucoside.

In conclusion, this paper reports the cDNA and amino acid sequence of the β-fucosidase/β-glucosidase, dalcochinin 8′-*O*-β-glucoside β-glucosidase, from the Thai rosewood, *D. cochinchinensis* Pierre. The amino acid sequence indicates that this protein belongs to the glycosyl hydrolase family 1 and is evolutionarily most closely related to cyanogenic β-glucosidases and other closely related β-glucosidases (*15*, *16*). Possible *N*-linked glycosylation sites were identified, and one site was confirmed by Edman degradation and mass spectroscopy. The activity of recombinant enzyme produced from the clone was shown to be essentially the same as β-glucosidase purified from seeds. Further investigation will be necessary to clarify the functional roles of various amino acids in the protein and the biological function of the protein.

The authors are greatly indebted to Dr. Rudee Surarit, who provided excellent advice and discussion throughout the project, and Dr. Chuenchit Boochird, who assisted in recombinant protein expression in *P. pastoris*. Phannee Sawangareetrakul, Pantipa Subhasitanont, Daranee Chokchaichamnankit, and Kanokporn Boonpuan provided technical assistance in various aspects of protein purification.

REFERENCES

1. Esen, A. (1993) β-Glucosidases overview in *β-Glucosidases Biochemistry and Molecular Biology* (Esen, A., ed.) pp. 1–14, American Chemical Society, Washington, DC
2. Conn, E.E. (1993) β-Glycosidases in plants: substrate specificity in *β-Glucosidases Biochemistry and Molecular Biology* (Esen, A., ed.) pp. 15–41, American Chemical Society, Washington, DC
3. Poulton, J.E. (1990) Cyanogenesis in plants. *Plant Physiol.* **94**, 401–405
4. Nisius, A. (1988) The stromacentre in *Avena plastids*: an aggregation of β-glucosidase responsible for the activation of oat-leaf saponins. *Planta* **173**, 474–481
5. Duroux, L., Delmotte, F.M., Lancelin, J.-M., Keravis, G., and Jay-Alleand, C. (1998) Insight into naphthoquinone metabolism: β-glucosdiase-catalysed hydrolysis of hydrojuglone β-D-glucopyranoside. *Biochem. J.* **333**, 275–283
6. Kuroki, G.W. and Poulton, J.E. (1987) Isolation and characterization of multiple forms of prunasin hydrolase from black cherry (*Prunus serotina* Ehrh.) seeds. *Arch. Biochem. Biophys.* **255**, 19–26
7. Henrissat, B. (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* **280**, 309–316
8. Brzobohaty, B., Moore, L., Kristoffersen, P., Bako, L., Campos, N., Schell, J., and Palmet, K. (1993) Release of active cytokinin by a β-glucosidase localized to the maize root meristem. *Science* **262**, 1051–1053
9. Malboobi, M.A. and Lefebvre, D,D. (1997) A phosphate-starvation inducible β-glucosidase gene (psr3.2) isolated from *Arabidopsis thaliana* is a member of a distinct subfamily of the BGA family. *Plant Mol. Biol.* **34**, 157–168
10. Nilsson, K. (1988) Enzymatic synthesis of oligosaccharides. *Trends Biotechnol.* **6**, 256–264
11. Welply, J.K. (1989) Sequencing methods for carbohydrates and their biological applications. *Trends Biotechnol.* **7**, 5–8

12. Surarit, R., Svasti, J., Srisomsap, C., Suginta, W., Khunyoshyeng, S., Nilwarangkoon, S., Harnsakul, P., and Benjavongkulchai, E. (1995) Screening of glycohydrolase enzymes in Thai plant seeds for potential use in oligosaccharide synthesis. *J. Sci. Soc. Thailand* **21**, 293–303
13. Srisomsap, C., Svasti, J., Surarit, R., Champattanachai, V., Sawangareetrakul, P., Boonpuan, K., Subhasitanont, P., and Chokchaichamnankit, D. (1996) Isolation and characterization of an enzyme with β-glucosidase and β-fucosidase activities from *Dalbergia cochinchinenesis* Pierre. *J. Biochem.* **119**, 585–590
14. Surarit, R., Chiba, S., Matsui, H., Svasti, J., and Srisomsap, C. (1996) Chemical modification of β-glucosidase/β-fucosidase from *Dalbergia cochinchinensis* Pierre by conduritol B epoxide. *Biosci. Biotechnol. Biochem.* **60**, 1265–1268
15. Svasti, J., Srisomsap, C., Techasakul, S., and Surarit, R. (1999) Dalcochinin-8′-*O*-D-glucoside and its β-glucosidase enzyme from *Dalbergia cochinchinensis*. *Phytochemistry* **50**, 739–743
16. Henrissat, B. and Bairoch, A. (1993) New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* **293**, 781–783
17. Chomczynski, P. and Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**, 156–159
18. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
19. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic Local Alignment Search Tool. *J. Mol. Biol.* **215**, 403–410
20. Feng, D.F. and Doolittle, R.F. (1996) Progressive alignment of amino acid sequences and construction of phylogenetic trees from them. *Methods Enzymol.* **266**, 368–382
21. Barrett, T., Suresh, C.G., Tolley, S.P., Dodson, E.J., and Hughes, M.A. (1995) The crystal structure of a cyanogenic β-glucosidase from white clover, a family 1 glycosyl hydrolase. *Structure* **3**, 951–960
22. Guex, N. and Peitsch, M.C. (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* **18**, 2714–2723
23. Hays, W.S., Jenison, S.A., Yamada, T., Pustuszyn, A., and Glew, R.H. (1996) Primary structure of the cytosolic β-glucosidase of guinea pig liver. *Biochem. J.* **319**, 329–837
24. Withers, S.G. and Street, I.P. (1988) Identification of a covalent α-D-glucopyranosyl enzyme intermediate formed on a β-glucosidase. *J. Am. Chem. Soc.* **110**, 8551–8553
25. Keresztessy, Z., Kiss, L., and Hughes, M.A. (1994) Investigation of the active site of the cyanogenic β-D-glucosidase (linamarase) from *Manihot esculenta* Crantz (cassava). II. Identification of Glu-198 as an active site carboxylate group with acid catalytic function. *Arch. Biochem. Biophys.* **315**, 323–330
26. Inoue, K., Shibuya, M., Yamamoto, K., and Ebizuka, Y. (1996) Molecular cloning and bacterial expression of a cDNA encoding furostanol glycoside 26-*O*-β-glucosidase of *Costus speciosis*. *FEBS Lett.* **389**, 273–277
27. Akiyama, T., Kaku, H., and Shibuyy, N. (1998) A cell wall-bound β-glucosidase from germinated rice: purification and properties. *Phytochemistry* **48**, 49–54
28. Burmeister, W.P., Cottaz, S., Driguez, H., Iori, R., Palmieri, S., and Henrissat, B. (1997) The crystal structures of *Sinapis alba* myrosinase and a covalent glycosyl-enzyme intermediate provide insights into the substrate recognition and active-site machinery of an *S*-glycosidase. *Structure* **5**, 663–675
29. Lin Cereghino, J. and Cregg, J.M. (2000) Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. *FEMS Microbiol. Rev.* **24**, 45–66
30. Keresztessy, Z., Hughes, M.A., and Liptak, A. (1996) Co-purification from *Eschericihia coli* of a plant β-glucosidase-glutathione S-transferase fusion protein and the bacterial chaperonin GroEL. *Biochem. J.* **314**, 41–47
31. Cicek, M. and Esen, A. (1999) Expression of soluble and catalytically active plant (monocot) β-glucosidases in *E. coli*. *Biotechnol. Bioeng.* **63**, 392–400
32. Cregg, J.M. (1999) Expression in the methylotrophic yeast *Pichia pastoris* in *Gene Expression Systems* (Fernandez, J. and Hoeffler, J., eds.) pp. 157–191, Academic Press, San Diego, CA